

Otros informes

Web semántica en bibliotecas: del material del que están hechos los sueños

Jesús Tramullas



Introducción

La web semántica, el paraíso soñado de **Tim Berners-Lee**, está tardando en llegar más de lo previsto. Las promesas del acceso a la información interrelacionada semántica y significativamente en tiempo real, mediante cualquier dispositivo, siguen siendo una

meta anhelada. Lo cierto es que, tras veinte años de World Wide Web, y diez del escrito seminal de **Berners-Lee, Hendler y Lassila** en *Scientific American*¹, resulta sorprendente que las tecnologías semánticas no hayan alcanzado las expectativas previstas. En 1988 **Berners-Lee** y el W3C ya habían publicado un mapa de ruta de la web semántica².

Las herramientas software necesarias para crear y explotar semánticamente conjuntos de información ya existen, tanto en productos comerciales, como en numerosas herramientas resultantes de proyectos de investigación. Se dispone de un amplio conjunto de lenguajes y esquemas de etiquetado de información, que permiten hacer interoperables grandes volúmenes de información... Con todo esto disponible, en algún lugar debe encontrarse el motivo o motivos del retraso.



Etiquetar semánticamente cuesta

La web se ha visto como un sistema de publicación textual rápida, sin especiales complicaciones técnicas, al alcance de cualquier ciudadano. La facilidad para crear documentos en html, unida a

la popularización de editores orientados a la edición –que no al etiquetado significativo de información–, trajo como consecuencia que durante una década se creara y publicara gran cantidad de información textual y gráfica, pero sin apenas procesado semántico útil para otros usos. ¿Durante cuánto tiempo se ha discutido sobre el interés que tenían los metadatos para usos avanzados, mientras se comprobaba que los creadores de información hacía caso omiso de su utilización? El problema era que etiquetar semánticamente los documentos generados requería un esfuerzo (y un coste) suplementario que muchos creadores de contenidos no estaban dispuestos a asumir. Y cuando decidían acudir a una referencia para copiar sus metadatos se encontraban con esquemas de clasificación de origen bibliotecario, de imposible aplicación a su contexto y necesidades, que además no ofrecían versiones etiquetadas reutilizables.

Ayuda de la web 2.0 y RDF

Este panorama cambió a comienzos de la década de 2001 gracias a la popularización de los servicios y aplicaciones de la web 2.0. Los gestores de contenidos más básicos, los blogs, requieren de sus usuarios que organicen la información en categorías, que la etiqueten..., y los usuarios empiezan a querer y a enlazar información relacionada con sus propios contenidos. Se establecen pautas intuitivas de enlazado semántico selectivo por parte de los usuarios finales, que aprovechan las funcionalidades a su alcance. Una mayor comprensión de la arquitectura y funcionamiento de los sistemas de gestión de contenidos y de la importancia de almacenar la información de manera estructurada en bases de datos trae una sensibilidad notable hacia la importancia de etiquetar, enlazar y reutilizar la información automáticamente, en virtud de su contenido semántico. La aparición –en los sistemas de gestión de contenidos más extendidos– de funciones capaces de crear e integrar información etiquetada en RDF³, junto al número creciente de recursos etiquetados en tripletas RDF, como muestra el desarrollo de *linked data*⁴ y los cada vez más numerosos *Sparql endpoints*⁵ están permitiendo que se llegue a un punto clave para la expansión de la web semántica en servicios y aplicaciones web.

Valor de los opacs

En este punto es necesario volver a las bibliotecas. Las bases de datos que sostienen los opacs contienen una de las principales fuentes de información estructurada disponible en el mundo. Sin

embargo, y a pesar de la importancia que los estándares tienen en la informatización bibliotecaria, lo cierto es que no están pensados para la reutilización de información, sino para la gestión bibliotecaria. El W3C es bien consciente de ello y del importantísimo papel que los datos bibliotecarios pueden jugar en el marco de *linked data*, y ha mantenido entre 2010 y 2011 el *W3C Library Linked Data Incubator Group*, cuya actividad, recomendaciones y resultados ha ido publicando y actualizando en la wiki correspondiente⁶.

Borrador del informe final⁷

Se publicó el 25 de agosto de 2011, y contiene claras y contundentes afirmaciones sobre las acciones que deben llevar a cabo las bibliotecas si quieren integrar sus conjunto de datos en el web semántica a través de *linked data*. No se trata de un informe largo, y su lectura completa no sólo es recomendable sino que es obligada. El 31 de agosto, *Library journal* se hizo eco de su publicación, dedicándole un comentario de **Michael Kelley** titulado *How the W3C has come to love library linked data*⁸.

Se estructura en ocho secciones, de las cuales son de especial relevancia las secciones 4 a 6, y los dos apéndices:

Sección 4⁹

Beneficios que tendría para los datos bibliotecarios la integración con y como *linked data*, en especial con referencia al enriquecimiento de la información, y a los procesos de descubrimiento de información y como *linked data*, en especial con referencia a la navegación y la recomendación.

Sección 5¹⁰

Situación actual de los datos bibliotecarios y sus silos, y no resulta muy halagüeña. Los datos bibliotecarios no se integran con otros silos, sus estándares son muy rígidos, pensados en y para comunidades cerradas, los identificadores están diseñados para un uso local en la mayoría de las ocasiones y no pueden enlazarse con otros contenidos, la calidad de los datos es muy variada...

Sección 6¹¹

Es la sección nuclear, y contiene un conjunto de recomendaciones. La clave estaría en hacer los datos bibliotecarios disponibles para su uso como *linked data*, al mismo tiempo que se usan *linked data* externos para enriquecer los servicios bibliotecarios. Es necesario identificar conjuntos de datos de valor que puedan publicarse dentro de proyectos asumibles, y adecuadamente dimensionados (no todo vale). De especial importancia es cambiar el enfoque en el desarrollo de estándares, atendiendo a la participación de otras comunidades de expertos en web semántica.

Técnicamente, la creación, gestión y preservación de URIs, así como el desarrollo y gestión de vocabularios en RDF son imprescindibles. La experiencia bibliotecaria en preservación digital es un factor a favor para la necesaria gestión de vocabularios en los próximos años.

Apéndices

Contienen un conjunto de referencias breves a diferentes recursos y herramientas. No deben pasar desapercibidos los enlaces a los entregables dedicados a *Casos de uso*¹² y a *Conjuntos de datos, vocabularios y conjuntos de elementos de metadatos*¹³, que ofrecen un rico panorama de propuestas, proyectos y recursos sobre el desarrollo de *linked data* en servicios bibliotecarios.

Algunas afirmaciones

Unas cuantas citas, en especial de títulos de apartados, que resultan significativas por sí mismas:

"Los datos bibliotecarios no están integrados con recursos web";

"La comunidad bibliotecaria y la comunidad semántica usan diferentes terminologías para conceptos similares de metadatos";

"Los estándares bibliotecarios se diseñan sólo para la comunidad bibliotecaria";

"Desarrollar estándares de datos bibliotecarios que sean compatibles con *linked data*";

"Diseñar historias de usuarios y modelos para interfaces de usuario";

"Conectar los estándares de datos bibliotecarios con los estándares de la web semántica";

Sin olvidar que "Los cambios en la tecnología bibliotecaria dependen de los desarrollos de los vendedores de sistemas". De nuevo hay que pensar en las soluciones de software libre.

Reflexión final

Los *linked data* y su evolución debería traer a la palestra una reflexión sobre la viabilidad de un modelo bibliotecario que se sigue desarrollando a expensas de propuestas y avances tecnológicos provenientes de otros contextos y comunidades. Si bien es cierto que existen notables y punteros proyectos y servicios bibliotecarios en la Web, no es menos cierto que las bibliotecas, en muchas ocasiones, adoptan un papel de adaptador de segunda o tercera ola. En un contexto científico, social y económico en el que se valora la innovación, esto no parece una estrategia inteligente. Realmente, este borrador de informe no dice nada que no sepamos o pueda considerarse novedoso; lo que puede preocupar es que todavía sea necesario decírselo a las bibliotecas.

Para ilustrar las posibilidades de los *linked data* en bibliotecas, es mejor terminar con el podcast

Register for free at <https://www.scipedia.com> to download the version without the watermark

de **Dan Brickley** para *The Harvard Library Innovation Laboratory* de la *Harvard Law School*¹⁴.

Notas

1. <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>
2. <http://www.w3.org/DesignIssues/Semantic.html>
3. <http://www.w3.org/RDF>
4. <http://linkeddata.org>
5. <http://www.w3.org/wiki/SparqlEndpoints>
6. http://www.w3.org/2005/Incubator/Ild/wiki/Main_Page#Library_Linked_Data_Incubator_Group_wiki
7. <http://www.w3.org/2005/Incubator/Ild/wiki/DraftReportWithTransclusion>
8. http://www.libraryjournal.com/lj/home/891826-264/how_the_w3c_has_come.html.csp
9. <http://www.w3.org/2005/Incubator/Ild/wiki/DraftReportWithTransclusion#Benefits>
10. <http://www.w3.org/2005/Incubator/Ild/wiki/DraftR>

[eportWithTransclusion#The_Current_Situation](http://www.w3.org/2005/Incubator/Ild/wiki/DraftReportWithTransclusion#The_Current_Situation)

11. <http://www.w3.org/2005/Incubator/Ild/wiki/DraftReportWithTransclusion#Recommendations>

12. <http://www.w3.org/2005/Incubator/Ild/wiki/UseCaseReport>

13. http://www.w3.org/2005/Incubator/Ild/wiki/Vocabulary_and_Dataset

14. <http://librarylab.law.harvard.edu/blog/2011/08/30/podcast-dan-brickley-on-libraries-linked-data-and-cataloguing-the-web>
tramullas@unizar.es

W3C Library Linked Data Incubator Group. Draft report with transclusion. Sept. 2011.

<http://www.w3.org/2005/Incubator/Ild/wiki/DraftReportWithTransclusion>

Nota original publicada en *Blok de BiD*:
<http://www.ub.edu/blokdebid/es/content/web-semántico-en-bibliotecas-del-material-del-que-están-hechos-los-sueños>

The logo for Scipedia, featuring a stylized yellow 'S' and 'C' that form a continuous shape, followed by the word 'IPEDIA' in a grey, sans-serif font.

SCIPEDIA

Register for free at <https://www.scipedia.com> to download the version without the watermark